



TITLE:

# A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing

AUTHOR(S):

Kato, Masaki; Takemoto, Keiko; Shinkai, Yoichi

---

CITATION:

Kato, Masaki ...[et al]. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. Nature Communications 2018, 9: 1683.

ISSUE DATE:

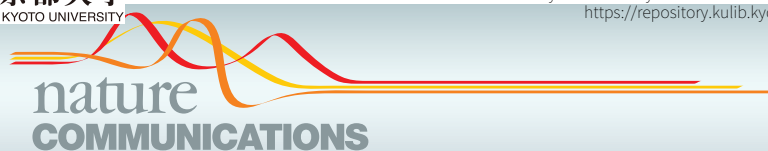
2018-04-27

URL:

<http://hdl.handle.net/2433/231147>

RIGHT:

© The Author(s) 2018. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



## ARTICLE

DOI: 10.1038/s41467-018-04132-9

OPEN

# A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing

Masaki Kato<sup>1</sup>, Keiko Takemoto<sup>2</sup> & Yoichi Shinkai<sup>1</sup>

Subsets of endogenous retroviruses (ERVs) are derepressed in mouse embryonic stem cells (mESCs) deficient for Setdb1, which catalyzes histone H3 lysine 9 trimethylation (H3K9me3). Most of those ERVs, including IAPs, remain silent if *Setdb1* is deleted in differentiated embryonic cells; however they are derepressed when deficient for *Dnmt1*, suggesting that Setdb1 is dispensable for ERV silencing in somatic cells. However, H3K9me3 enrichment on ERVs is maintained in differentiated cells and is mostly diminished in mouse embryonic fibroblasts (MEFs) lacking Setdb1. Here we find that distinctive sets of ERVs are reactivated in different types of *Setdb1*-deficient somatic cells, including the VL30-class of ERVs in MEFs, whose derepression is dependent on cell-type-specific transcription factors (TFs). These data suggest a more general role for Setdb1 in ERV silencing, which provides an additional layer of epigenetic silencing through the H3K9me3 modification.

<sup>1</sup> Cellular Memory Laboratory, Cluster for Pioneering Research, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. <sup>2</sup> Institute for Frontier Life and Medical Sciences, Kyoto University, 53 Shogoin, Sakyo, Kyoto 606-8507, Japan. These authors contributed equally: Masaki Kato, Keiko Takemoto. Correspondence and requests for materials should be addressed to M.K. (email: [mkato@riken.jp](mailto:mkato@riken.jp)) or to Y.S. (email: [yshinkai@riken.jp](mailto:yshinkai@riken.jp))

Endogenous retroviruses (ERVs) are major components of mammalian genomes<sup>1</sup>. To silence ERV activity and transposition, hosts have evolved multiple silencing mechanisms to establish and maintain repressive heterochromatin formation across these elements<sup>2</sup>. We previously demonstrated that the H3K9me3 methyltransferase *Setdb1* (Eset, KMT1E) represses intracisternal A particles (IAPs) and other ERVs in mouse embryonic stem cells (mESCs)<sup>3,4</sup>. Similarly, the depletion of Trim28 (Kap1, Tif1 $\beta$ ), which is a co-repressor of *Setdb1*, leads to the derepression of IAP elements in mESCs<sup>5</sup>. Trim28 is recruited to ERVs via sequence-specific Krüppel-associated box zinc-finger proteins (KRAB-ZFPs), a large family of transcription factors (TFs) that might have expanded and co-evolved in mammalian genomes with the expansion of ERVs<sup>6,7</sup>. DNA methylation is mostly dispensable for the silencing of ERVs repressed by *Setdb1* in early embryonic cells<sup>3,4</sup>. *Setdb1* is also required for the silencing of ERVs at approximately embryonic day 13.5 (E13.5) in primordial germ cells<sup>8</sup>, in which DNA is typically hypomethylated<sup>9,10</sup>. In contrast, in somatic and male germ-lineage cells, which are further differentiated, DNA methylation has been proposed to be the main mechanism to silence ERVs and non-long terminal repeat (LTR) retroelements. In this context, CpG dinucleotides of retroelements are densely methylated and *Dnmt1* inactivation leads to IAP derepression in embryos or differentiated mESCs upon depletion of LIF signals<sup>11,12</sup>. Thus, it is generally considered that the H3K9me3-mediated ERV silencing pathway in mESCs is rapidly replaced by a more permanent silencing mechanism, i.e., Trim28-mediated de novo DNA methylation, in differentiated embryonic cells<sup>13–15</sup>. Once DNA methylation is established, sequence-specific KRAB-ZFPs and Trim28 are no longer required<sup>16,17</sup>.

DNA methylation patterns acquired during development have long been considered a stable epigenetic mark in somatic cells and adult cells. However, several recent studies have revealed that some ERVs are also derepressed in differentiated somatic cells lacking Trim28 or *Setdb1*<sup>18–22</sup>. In this study, we re-evaluate the role of *Setdb1* in ERV silencing, not only in ESCs, but also in differentiated somatic cells, in which ERVs are heavily DNA methylated. We find that specific sets of ERVs are reactivated in different types of *Setdb1*-deficient somatic cells. Our data suggest that *Setdb1* plays a more general role in ERV silencing, providing an additional silencing mechanism through H3K9me3.

## Results

**Derepression of distinct ERV families upon loss of *Setdb1*.** H3K9me3 enrichment in ERV family members has been detected in mESCs<sup>3–5,23,24</sup>. Although DNA methylation is important for ERV silencing in differentiated cells, the relevance of H3K9me3 marks is not well defined. To analyze whether the H3K9me3 marks on ERVs are important for silencing in differentiated cell types, we performed RNA sequencing (RNA-seq) analysis on *Setdb1* conditional knockout (cKO) immortalized mouse embryonic fibroblasts (iMEFs)<sup>3</sup>, which is a model for differentiated cells. Our data were compared with previously published RNA-seq datasets for mESCs and other differentiated cell types with or without *Setdb1* KO<sup>4,18,20</sup>. The amount of *Setdb1* in iMEFs is almost 10 times lower than that in ESCs, and depletion of *Setdb1* by 4-hydroxytamoxifen (4OHT) in iMEFs induced growth defects, similar to that in *Setdb1* cKO mESCs. However, the growth recovered 8 days after 4OHT treatment in iMEFs was unlike that in ESCs<sup>3</sup> (Supplementary Fig. 1). We analyzed total RNA (rRNA was depleted) from untreated and 4OHT-treated *Setdb1* cKO iMEFs 5 days after treatment. An RNA-seq analysis of repeats in *Setdb1* cKO iMEFs revealed a substantially increased expression of ERVs after *Setdb1* depletion, particularly five

elements annotated by Repbase, i.e., MMVL30-int, MuLV-int, RLTR4\_Mm, RLTR4\_Mm-int, and RLTR6\_Mm (Fig. 1a, highlighted in red). In contrast, distinct ERV families were derepressed in other cell types when *Setdb1* was removed<sup>4,18–20,25</sup>. For example, MMERVK10C exhibited the highest induction in *Setdb1* KO ESCs. Furthermore, IAPLTR1\_1a\_ and 2\_Mm were highly derepressed in the fetal forebrains of *Setdb1* KO mice, and RLTR3\_Mm was specifically induced in *Setdb1* KO granulocyte/macrophage progenitors (GMPs) or bone marrow Lin<sup>–</sup> Sca-1<sup>+</sup> c-Kit<sup>+</sup> (LSK) cells.

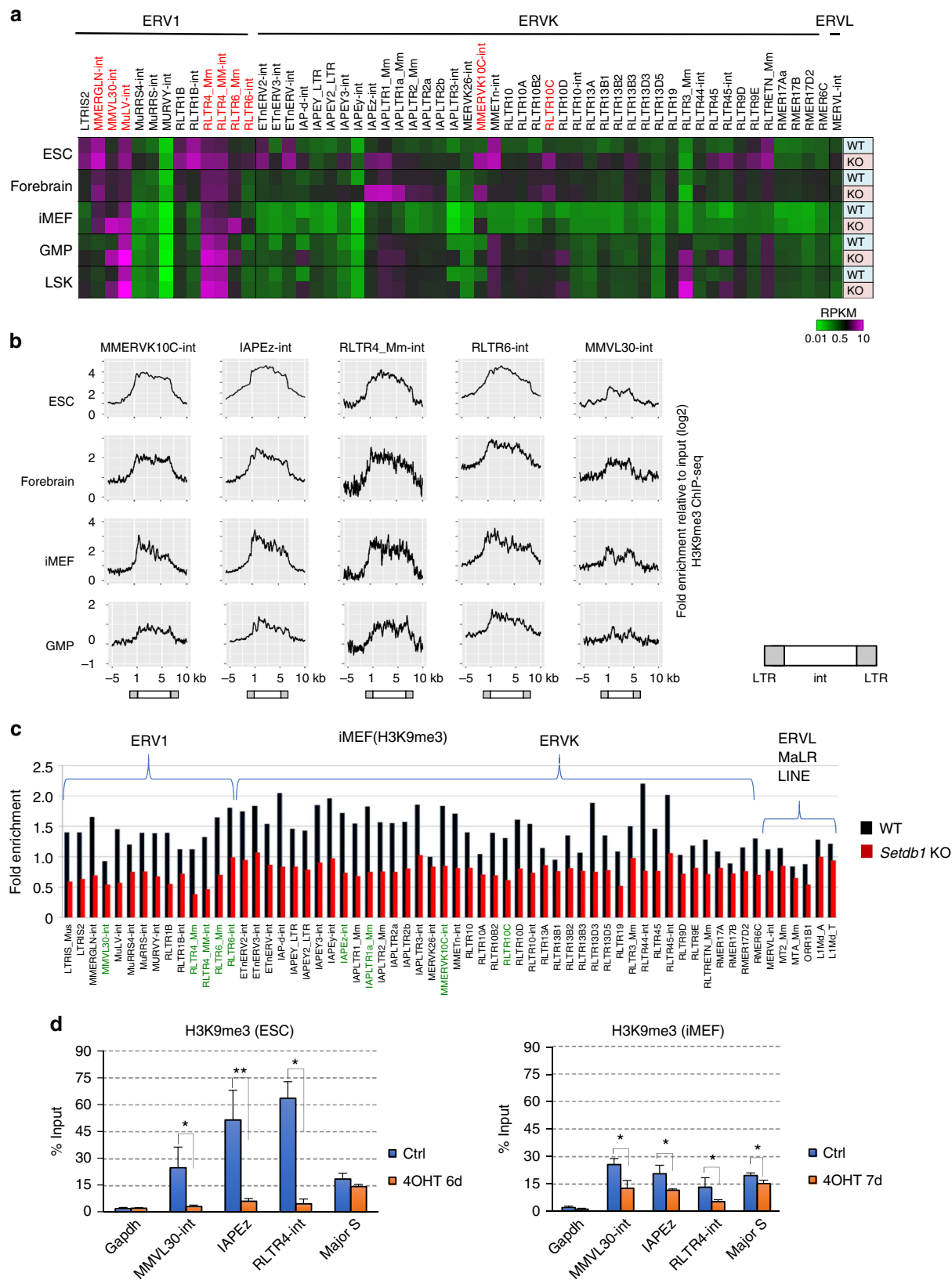
**A viral defense response is induced in *Setdb1* KO iMEFs.** We also examined the transcription of non-repeats, and identified 244 RefSeq genes that were upregulated by at least 2-fold in 4OHT-treated *Setdb1* cKO iMEFs, compared to untreated cKO iMEFs (Supplementary Fig. 2a and Supplementary Data 1). A Gene Ontology analysis of the upregulated genes revealed the major representation of genes involved in immune or defense responses (Supplementary Fig. 2b). In particular, we observed the upregulation of several genes related to interferon (IFN) responses, such as *Irf7*, *Irf9*, *Usp18*, and *Stat1* in *Setdb1* KO iMEFs (Supplementary Fig. 2c). *Irf7*, which is a master regulator of type-I IFN-dependent immune responses, was upregulated by more than 20-fold.

To determine how many genes upregulated in *Setdb1* KO iMEFs were directly repressed by *Setdb1*, we performed a chromatin immunoprecipitation sequencing (ChIP-seq) analysis of H3K9me3 in wild-type (WT) iMEFs. Surprisingly, only 1% of the promoter regions of genes that were upregulated in *Setdb1* KO iMEFs (2/244) were marked by H3K9me3 (Supplementary Fig. 3a), indicating that almost all genes induced by *Setdb1* KO are indirectly regulated by *Setdb1*. This is consistent with the results obtained for *Setdb1* cKO mESCs<sup>4</sup>. Derepressed ERVs might play a role in altering the gene expression profile of *Setdb1* KO cells, since transcriptionally reactivated ERVs significantly enhance the transcription of their neighboring genes<sup>4,18</sup>. Thus, we analyzed the correlation between upregulated genes in *Setdb1* KO iMEF cells and ERV insertions. In MEFs, ~50% of the H3K9me3-marked upregulated genes (11 out of 27 genes) had ERV or LINE insertions with H3K9me3 enrichment (Supplementary Fig. 3a). H3K9me3 on these retroelements diminished in MEFs after *Setdb1* depletion (Supplementary Fig. 3b and Supplementary Data 2). However, we did not observe any derepressed elements inserted in the upregulated genes in *Setdb1* KO iMEFs, unlike those seen in *Setdb1* KO ESCs or forebrain cells<sup>4,18</sup>. Furthermore, we calculated the ratio between upregulated genes with and without ERV insertions, and compared it with those for the downregulated and unchanged genes. There was no statistically significant difference between the values for the upregulated genes and those for the other genes (the ratio of upregulated genes is 0.0472, downregulated 144 genes 0.0434, and unchanged 22533 genes 0.0467, respectively, thus ERVs seen at upregulated genes were expected by chance ( $P = 0.12$ , binomial test)).

**H3K9me3 on ERVs are preserved in distinct cell types.** Next, we tried to determine whether H3K9me3 marks on ERVs seen in ESCs are preserved in distinct differentiated somatic cells. For this, we utilized published H3K9 ChIP-seq datasets for ESCs<sup>24</sup>, fetal forebrain cells<sup>18</sup>, iMEFs<sup>24</sup>, and GMPs<sup>20</sup>, and generated NGS plots using the positions of H3K9me3 peaks for representative ERVs (MMERVK10C-int, IAPez-int, RLTR4\_Mm-int, RLTR6-int, and MMVL30-int), which are all derepressed in at least one of the studied *Setdb1* KO cell types (Fig. 1b). In this analysis, we selected full-length ERV elements for each family. Comparison of

H3K9me3 profiles demonstrated that the pattern of H3K9me3 peaks on ERVs are chiefly conserved among these distinct cell types, although H3K9me3 enrichment seems lower in differentiated cell types than in ESCs, which is consistent with the results from a previous report<sup>24</sup>.

**Setdb1-dependent H3K9me3 marks on ERVs.** To address why *Setdb1* KO led to the derepression of distinct sets of ERV families, we performed ChIP-seq analysis of H3K9me3 on *Setdb1* cKO iMEFs treated with or without 4OHT (KO) for 7 days, and compared the data for the WT and KO samples. Genome-wide



ChIP-seq analysis revealed that most of the enriched H3K9me3 on the examined ERVs was reduced in *Setdb1* cKO iMEFs due to 4OHT treatment; however, the reduction level on each ERV was variable (Fig. 1c and Supplementary Data 2). We observed a slight reduction of H3K9me3 in LINE families (Fig. 1c and Supplementary Fig. 3c). To validate the ChIP-seq data and to compare *Setdb1*-dependent H3K9me3 on ERVs between iMEFs and ESCs, we performed a ChIP-qPCR analysis of H3K9me3 on selected ERVs and major satellite repeats for both cell types (Fig. 1d). Consistent with previous results and NGS plot data (Fig. 1b), H3K9me3 was enriched in the repeat domains of WT ESCs, but exhibited much less enrichment on such repeats in iMEFs. After *Setdb1* depletion in *Setdb1* cKO mESCs and iMEFs, we observed a decrease in H3K9me3 enrichment at the MMVL30-int, IAPez, and RLTR4-int loci, suggesting that those H3K9me3 modifications were induced by *Setdb1* in iMEFs (Fig. 1d). In contrast, we did not observe any significant reduction of H3K9me3 at major satellite loci in ESCs; however, a slight reduction was observed in iMEFs, where *Suv39h* is responsible for most of H3K9me3<sup>24</sup>. Thus, *Setdb1* regulated H3K9me3 marks at many of the ERV loci that were examined.

**Derepression of the VL30 family in *Setdb1* KO MEFs.** RNA-seq analysis of repeats in *Setdb1* cKO iMEFs revealed a substantial increase in the expression of MMVL30-int and RLTR6\_Mm sequences (Fig. 1a). RLTR6\_Mm is an LTR element flanking MMVL30-int in the VL30 family (virus-like 30, MMVL30). The VL30 ERV family can be largely classified into two groups, based on their internal sequences. One group has RLTR6-int elements, which are predominantly intact ERVs with coding regions for retroviral Gag, Pol, and Env proteins; the second group has MMVL30-int elements, which are likely derived from RLTR6-int elements, but have lost their coding regions (Fig. 2a). In addition, VL30-containing RLTR6 LTRs can be divided into four subgroups on the basis of their LTR U3 sequences<sup>26</sup>. RLTR6-int type VL30 is mostly flanked by U3 class II RLTR6s; therefore, we called it VL30 U3 class II. MMVL30-int is flanked by U3 class I, III, or IV RLTR6s, and were thus called VL30 U3 class I, III, or IV, respectively (Fig. 2a). U3 class III RLTR6s contain a DR2-type all-trans retinoic acid (atRA)-response element (RARE), and U3 class III VL30 expression is induced by atRA in keratinocytes<sup>27</sup>. On the other hand, U3 class II RLTR6s contain DR2-type and DR5-type RAREs (Fig. 2a).

To identify which VL30 U3 classes are repressed by *Setdb1* in iMEFs, we analyzed the RNA-seq data of iMEFs with and without *Setdb1* depletion by treatment with 4OHT. After 4OHT treatment, VL30 U3 class I and III were significantly derepressed by more than 10-fold and 5-fold, respectively, and class IV was marginally induced, but class II was unaffected (Fig. 2b). We generated NGS plots, using the positions of H3K9me3 peaks on

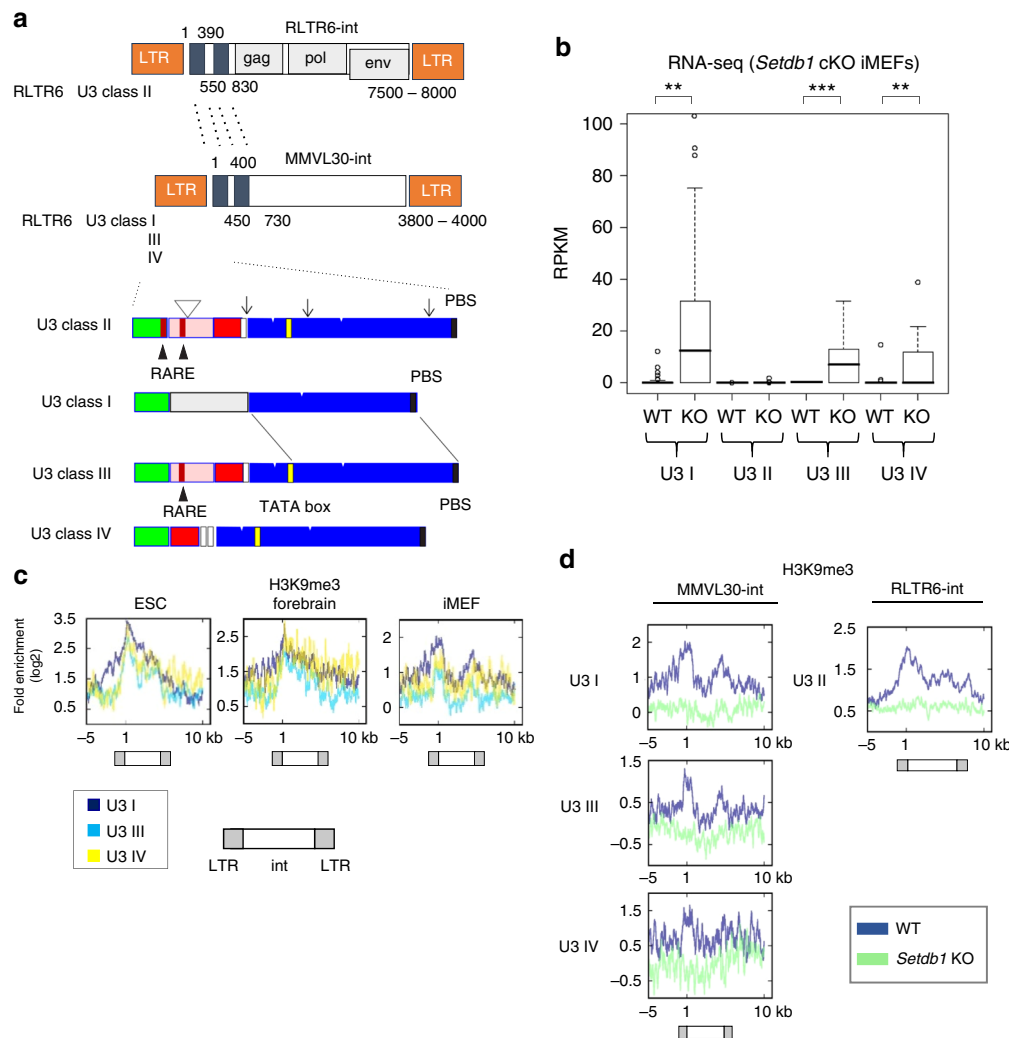
VL30 U3 class I, III, and VI elements containing the MMVL30-int sequence (Fig. 2c). They shared similar H3K9me3 peak patterns in the three cell types (ESCs, forebrain cells, and iMEFs), even though significant derepression of VL30 U3 class I and III was only observed in *Setdb1* KO iMEFs. We also observed similar H3K9me3 peak patterns on U3 class II-containing RLTR6-int sequences in the three cell types (Fig. 1b). ChIP-seq analysis of *Setdb1* KO iMEFs confirmed the loss of H3K9me3 in all four classes of VL30 (Fig. 2d).

**Distinct regulation mechanism for each VL30 U3 class.** Since specific U3 classes of VL30 were reactivated by *Setdb1* KO in iMEFs, we expected that each class would have different regulation mechanisms for transcription. VL30 U3 class II has a predominantly intact internal sequence (Rebase annotation: RLTR6-int) with coding regions for retroviral Gag, Pol, and Env proteins (Fig. 2a). The H3K9me3 peaks are enriched not only on the 5' LTR, but also across the internal region (Fig. 2d right). As stated, this VL30 U3 class II expression is restricted to Rar  $\alpha$  activity on LTR RAREs in the liver<sup>28</sup>. Therefore, the lack of derepression of VL30 U3 class II in *Setdb1* KO iMEFs might be due to an absence of endogenous atRA in MEFs. We then evaluated the requirement of atRA for VL30 U3 class II transcription. The addition of atRA alone did not activate VL30 U3 class II expression in *Setdb1* cKO iMEFs, but substantial induction was observed after both atRA and 4OHT treatments (Fig. 3a). In mESCs, *Setdb1* KO or atRA treatment alone weakly activated VL30 U3 class II; this activation was further enhanced after both atRA and 4OHT treatment (Fig. 3b). Thus, for the U3 class II VL30 expression in both MEFs and mESCs to be maximum, a loss of H3K9me3 and the presence of atRA is required. We also examined the effect of atRA on VL30 class I copies, but impact of atRA addition to *Setdb1* KO is marginal or absent for iMEFs and ESCs, respectively.

Next, we focused on distinct individual elements of VL30 U3 class I (total 71 elements) to further investigate interactions between *Setdb1*-mediated H3K9me3 and transcriptional silencing. Twenty-two elements have a typical proline tRNA primer-binding site (PBS-pro) near the 3' end of their 5' LTR. Not all of these are derepressed in *Setdb1* KO iMEFs (Fig. 3c top); PBS-pro-negative elements are relatively more derepressed than their PBS-pro-positive counterparts. Although unique mapping was incomplete due to highly homologous sequences, we tried to analyze the reduction of H3K9me3 at the 5' LTR regions of 71 VL30 U3 class I elements in *Setdb1* KO iMEFs using ChIP-seq data. We observed a significant reduction of H3K9me3 at most of those distinct element loci, but the reduction rate varied (Fig. 3c bottom). We observed a marginal correlation between the derepression rate and reduction rate of H3K9me3 for those VL30 class I elements in *Setdb1* KO iMEFs (Supplementary

**Fig. 1** Different ERV families are derepressed by *Setdb1* KO in different cell types. **a** Cell-type-specific ERV derepression in *Setdb1* cKO cells. Expression of ERV families in *Setdb1* cKO ESC (day 6 after treatment with 4OHT (KO) or no treatment (WT))<sup>4</sup>, iMEF (day 5 after treatment with 4OHT (KO) or no treatment (WT)), and E14.5 forebrain cells from WT and *Emx2-Cre:Setdb1* fl/fl mice (KO)<sup>18</sup>. For GMP and LSK cells, bone marrow cells from *Rosa-CreERT: Setdb1* cKO mice were transplanted into irradiated recipient mice, GMP and LSK cells were then isolated after injection of 4OHT for 2 weeks (KO) or control injection (WT)<sup>20</sup>. Only ERVs derepressed ( $\geq 2$  fold) in at least one of the analyzed cell types with *Setdb1* KO are listed. Heatmap indicates the relative expression level of representative ERV families (the RPKM value). The ERVs derepressed ( $\geq 1.5$  fold) in *Setdb1* KO iMEFs are highlighted in red. **b** H3K9me3 intensity profiles on different ERV families in different cell types. NGS plots show the fold enrichment of H3K9me3 from -5 kb to 10 kb around genomic ERV elements in ESC, forebrain, iMEF, and GMP. We selected ERVs containing -int element with flanked LTRs (See Methods). Position 1 is 5' start site of the -int element. Positions of LTRs and int for each ERV element are indicated below the plots. **c** Bar plots showing the loss of H3K9me3 in ERV families in *Setdb1* KO (day 7 after 4OHT treatment: red bar) vs. WT (black bar) iMEF. The y axis indicates the fold enrichment of normalized ChIP read density relative to input. ERV names written in green are analyzed in Fig. 1b. **d** ChIP-qPCR of H3K9me3 in the *Gapdh* promoter region indicated ERVs and major satellite (Major S) loci of *Setdb1* cKO ESCs and iMEFs. Values are means  $\pm$  s.d. from independent experiments ( $n = 3$ ). \*0.005 <  $P$  < 0.05, \*\*0.0005 <  $P$  < 0.005, Student's  $t$ -test





**Fig. 2** MMVL30 families are derepressed in *Setdb1* KO iMEFs. **a** Schematic structures of VL30 families. Blue boxes indicate highly homologous sequences between MMVL30-int and RLTR6-int. High magnification of LTR regions of U3 classes. Color boxes indicate highly homologous sequence among LTRs. White arrowhead and arrow indicate unique insertion and mutation in class II, respectively. Black arrowhead indicates RARE sites. **b** Expression of VL30 U3 classes in *Setdb1* KO iMEFs. Boxplots show that expression of VL30 U3 classes in *Setdb1* KO iMEFs (day 5 after 4OHT treatment (KO) or no treatment (WT)). RNA-seq reads overlapping each VL30 loci are counted and normalized using the length of the locus and million mapped reads (RPKM). U3 I (71 loci), U3 II (189 loci), U3 III (66 loci), and U3 IV (26 loci). \*\*0.0005 < P < 0.005, \*\*\*P < 0.0005, Student's t-test. **c** NGS plots of H3K9me3 ChIP-seq data at genomic VL30 loci (RLTR6\_Mm - MMVL30-int) of ESCs, E14.5 forebrain cells and iMEFs. MMVL30-int with RLTR6\_Mm U3 I (36), with U3 III (40) and with U3 IV (10). The number inside of parenthesis indicates the number of individual each element analyzed. LTRs are aligned and MMVL30-int starts from position 1. H3K9me3 intensity profiles are shown with dark blue line (U3 I), light blue one (U3 III), and yellow one (U3 IV). Positions of LTRs and int are indicated below the plots. **d** NGS plots of H3K9me3 ChIP-seq data at genomic "RLTR6\_Mm - MMVL30-int" and "RLTR6\_Mm - RLTR6-int" loci in *Setdb1* cKO iMEFs (day7 after 4OHT treatment (KO) and no treatment (WT)). H3K9me3 fold enrichments (log2) are shown with dark blue line (WT) and light blue one (KO).

Fig. 4). Since only a subset of VL30 U3 class I was derepressed in *Setdb1* KO cells, this derepression variation cannot be explained just by the reduction rate of H3K9me3.

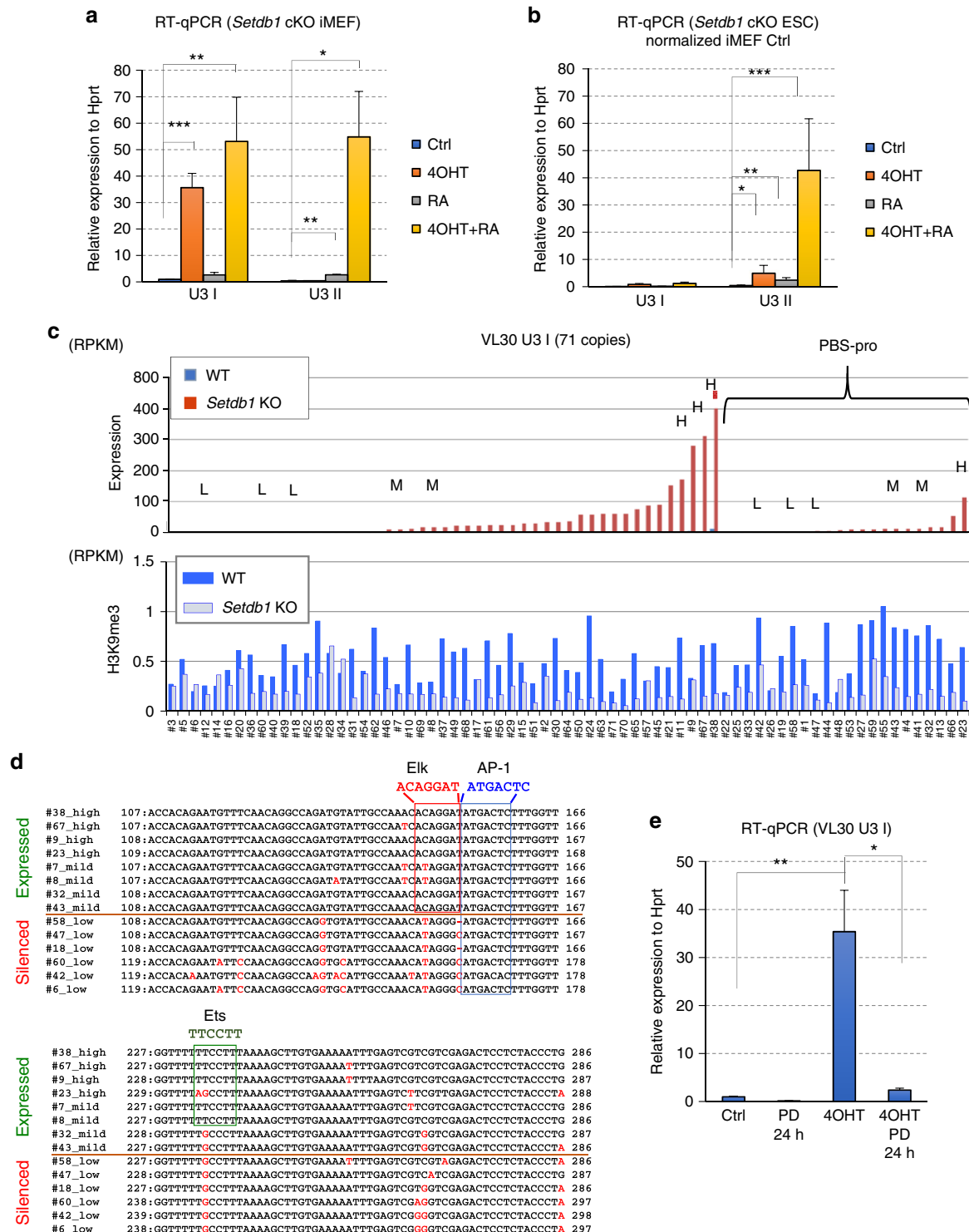
The structural analysis of LTRs of VL30 has been reported to show a possible requirement of tissue-specific TFs<sup>29</sup>. Thus, we focused on the potential role of TFs in VL30 U3 class I derepression in *Setdb1* KO iMEFs. Multiple DNA sequence alignment of the LTRs for VL30 U3 class I, generated using Clustal Omega, revealed that the silent populations differ slightly at the DNA sequence level (Fig. 3d). A motif analysis using the Patch 1.0 database predicted binding sites for several TFs, including AP-1, Ets1, and Elk, in the LTRs for VL30 U3 class I, but motifs for Elk family and Ets1-binding sites were mutated in the silent or low derepressed populations (Fig. 3d and Supplementary Data 3). Elk-1 is a member of the Ets family of

TFs. Elk-1 and Ets1 appear to be direct targets of activated MAP kinase (MAPK)<sup>30</sup>. Thus, we examined the requirement of MAPK activity for VL30 U3 class I derepression in *Setdb1* KO iMEFs. PD0325901 is a potent MEK inhibitor that suppresses the phosphorylation of ERK<sup>31</sup>. We treated *Setdb1* cKO iMEFs with PD0325901 for the final 24 h of a 5-day treatment with 4OHT and observed a strong inhibition of VL30 derepression in PD0325901-treated iMEFs (Fig. 3e). RNA-seq analysis showed that induction of low derepressed VL30 elements with mutated Elk/Ets binding sites, such as #58 and #47 copies were also diminished by the MEK inhibitor treatment as similar to those containing intact Elk/Ets binding sites (Supplementary Fig. 5a). Indeed, AP-1 is known to be present downstream of the MAPK pathway to activate VL30 elements<sup>32,33</sup> and AP-1 binding site is mostly intact in the derepressed VL30 elements regardless of Elk/

Ets binding site mutation(s) (Fig. 3d and Supplementary Data 3). Thus, we speculate that AP-1 also contribute to activation of VL30 U3 class I elements after Setdb1 depletion. To test the direct involvement of Ets1 in VL30 derepression, we performed ChIP-qPCR analysis of Flag-tagged Ets1 to see an enrichment of Ets1 at VL30 loci. We observed a slight enrichment of Ets1 at U3 class I #9 loci (Ets binding site is intact) after Setdb1 depletion, but no enrichment at #6 and #18 copies, in which the Ets-binding site is mutated (Supplementary Fig. 5b), suggesting that Ets1 might contribute to VL30 derepression. However, the Ets family is large, and we do not rule out the possibility that other Ets proteins are

involved in VL30 derepression. Considering these results, we conclude that most VL30-class ERVs are silenced by the Setdb1 pathway in MEFs. However, only specific sets of ERVs are potentially active, depending on whether the regulatory elements in their LTRs are compatible with the activity of TFs in distinct cell types.

**Distinct requirements for epigenetic marks in ERV silencing.** DNA methylation has an important role in proviral silencing in somatic cells<sup>12,34</sup>, but the impact of DNA methylation over entire



ERVs has not been well characterized, especially in differentiated cells. Therefore, we performed RNA-seq analysis on *Dnmt1* knock down (KD) iMEFs using siRNA. Surprisingly, only a small fraction of ERV families was derepressed in *Dnmt1* KD iMEFs, involving mostly the IAPEz family and the flanking LTR (IAPLTR1\_Mm), with a much weaker fold induction of MuLV-int and MMVL30-int (Fig. 4a top). We also reanalyzed public RNA-seq data for WT and *Dnmt1*<sup>-/-</sup> iMEFs in the p53<sup>-/-</sup> background<sup>35</sup> to examine the derepression of ERVs. Consistent with the *Dnmt1* KD result, we observed a prominent derepression of the IAPEz family and the flanking LTR (IAPLTR1\_Mm) (Supplementary Fig. 6). In *Dnmt1* KO ESCs, the ERVs IAPEz and IAPLTR1\_Mm were also prominently derepressed (Fig. 4a bottom)<sup>36</sup>.

Since DNA methylation is involved in silencing a subset of ERVs, such as IAPEz, in mESCs in the absence of H3K9me3<sup>4</sup>, we examined whether the simultaneous depletion of DNA methylation and *Setdb1* leads to a higher level of ERV reactivation in iMEFs than in the *Setdb1* KO alone. *Setdb1* cKO iMEFs were treated with siRNA specific to *Dnmt1* alone or in combination with 4OHT, and then RNA-seq analysis was performed. We did not observe a strong augmentation of ERV derepression by the simultaneous depletion of *Dnmt1* and *Setdb1* (Fig. 4a top). In contrast to VL30, IAPs possess the highest CpG density<sup>36</sup>. We also confirmed the RNA-seq data using RT-qPCR. Consistent with the *Dnmt1* KD iMEF and *Dnmt1*<sup>-/-</sup> iMEF RNA-seq results, *Dnmt1* KD alone induced a strong reactivation of IAPEz in iMEFs (Fig. 4b), but the *Setdb1* KO alone did not show any reactivation. IAPEz did not show increased reactivation by simultaneous depletion of *Dnmt1* and *Setdb1*. On the other hand, we observed a slight derepression of VL30 by *Dnmt1* KD alone. In addition, VL30 showed a minor augmentation in reactivation following the simultaneous depletion of *Dnmt1* and *Setdb1*, compared to that by *Setdb1* KO alone (Fig. 4b). For some of ERVs like RLTR4, we observed a mild derepression and marginal augmentation in induction due to the simultaneous depletion of both *Setdb1* and *Dnmt1* (Fig. 4b).

We further analyzed distinct individual elements of VL30 U3 class I to investigate the role of DNA methylation and H3K9me3 in their silencing (Fig. 4c). Some of the *Setdb1*-repressed elements like #9 and #38 (all non-PBS-pro) were also derepressed by *Dnmt1* KD with *Dnmt1*-specific siRNA, but most of these U3 class I elements were not derepressed, and the impact of the simultaneous depletion of *Dnmt1* and *Setdb1* was marginal (Fig. 4c). *Dnmt1* siRNA treatment significantly reduced CG methylation levels on IAPEz (at the Mnd1 locus) and VL30 (PBS-pro) (Supplementary Fig. 7a); however, we did not detect clear reduction in DNA methylation at the analyzed VL30 5' LTR sequences (individual locus of non-PBS-pro copies, #21, #38 and pool of PBS-pro elements loci) after the depletion of *Setdb1*

(Supplementary Fig. 7b and 8). Collectively, the silencing of VL30 in MEFs mostly depends on the *Setdb1* pathway, even though DNA methylation contributes to this phenomenon to some extent. In contrast, the *Dnmt1*-mediated pathway is critical for the repression of IAPEz in iMEFs. Both the *Dnmt1* and *Setdb1* pathways contribute to IAPEz silencing in mESCs<sup>4</sup>.

**Long-term-cultured *Setdb1* KO iMEFs.** In our RNA-seq analysis using long-term-cultured *Setdb1* KO iMEFs, we noticed that reactivated ERV expression was lower than that in 4OHT-treated *Setdb1* cKO iMEFs (Supplementary Fig. 9a). To further examine the dynamics of VL30 U3 class I derepression after the acute depletion of *Setdb1*, we performed RT-qPCR analysis of VL30 U3 class I from day 0 to 11 days after the addition of 4OHT (Supplementary Fig. 9b). Upregulation of VL30 U3 class I was observed from day 4, and by day 8, the transcription level was more than 60-fold higher than the WT levels. However, this VL30 U3 class I expression decreased progressively after day 9. We performed ChIP-qPCR for H3K9me3 in WT, 4OHT-treated *Setdb1* cKO iMEFs, and long-term-cultured *Setdb1* KO iMEFs, and observed a lower enrichment of H3K9me3 at VL30 loci in long-term-cultured *Setdb1* KO iMEFs, which was not the case in WT cells (Supplementary Fig. 9c). We did not observe any recovery of H3K9me3 after long-term culturing. It is possible that other epigenetic repressive modifications, like DNA methylation or H3K27me3 modification, function to silence VL30, as backup mechanisms. Additional treatment of long-term-cultured *Setdb1* KO iMEFs with 5-Aza-dC, a *Dnmt* inhibitor, did not increase VL30 expression (Supplementary Fig. 9d). Thus, the desensitization of VL30 expression in long-term-cultured *Setdb1* KO iMEFs is not due to DNA methylation as a backup mechanism. To examine the potential role of H3K27me3, we treated cells with GSK126, which is a specific inhibitor of Ezh2<sup>37</sup>. Depletion of H3K27me3 in long-term-cultured *Setdb1* KO iMEFs did not augment the expression of VL30 induced by *Setdb1* KO alone; however, the overall H3K27me3 level was significantly reduced (Supplementary Fig. 9e, f).

**Distinct requirements for *Setdb1*, *Trim28*, and *Zfp809*.** *Zfp809*, a member of the KRAB-ZFP family, initiates the silencing of VL30 family members in a sequence-specific manner via the recruitment of *Trim28*-*Setdb1* complexes<sup>38</sup>. *Zfp809* binds to PBS-pro, which is used by some retroviruses to prime reverse transcription. Thus, *Zfp809* can recruit *Trim28* and *Setdb1* to ERVs possessing PBS-pro<sup>39–41</sup>. We analyzed the RNA-seq data of *Setdb1* KO iMEFs and *Zfp809* KO MEFs<sup>38</sup> to compare non-PBS-pro and PBS-pro, with respect to VL30 derepression. The *Zfp809* KO MEFs were obtained from the KO embryo<sup>38</sup>. We observed a significant derepression of the non-PBS-pro VL30 group in

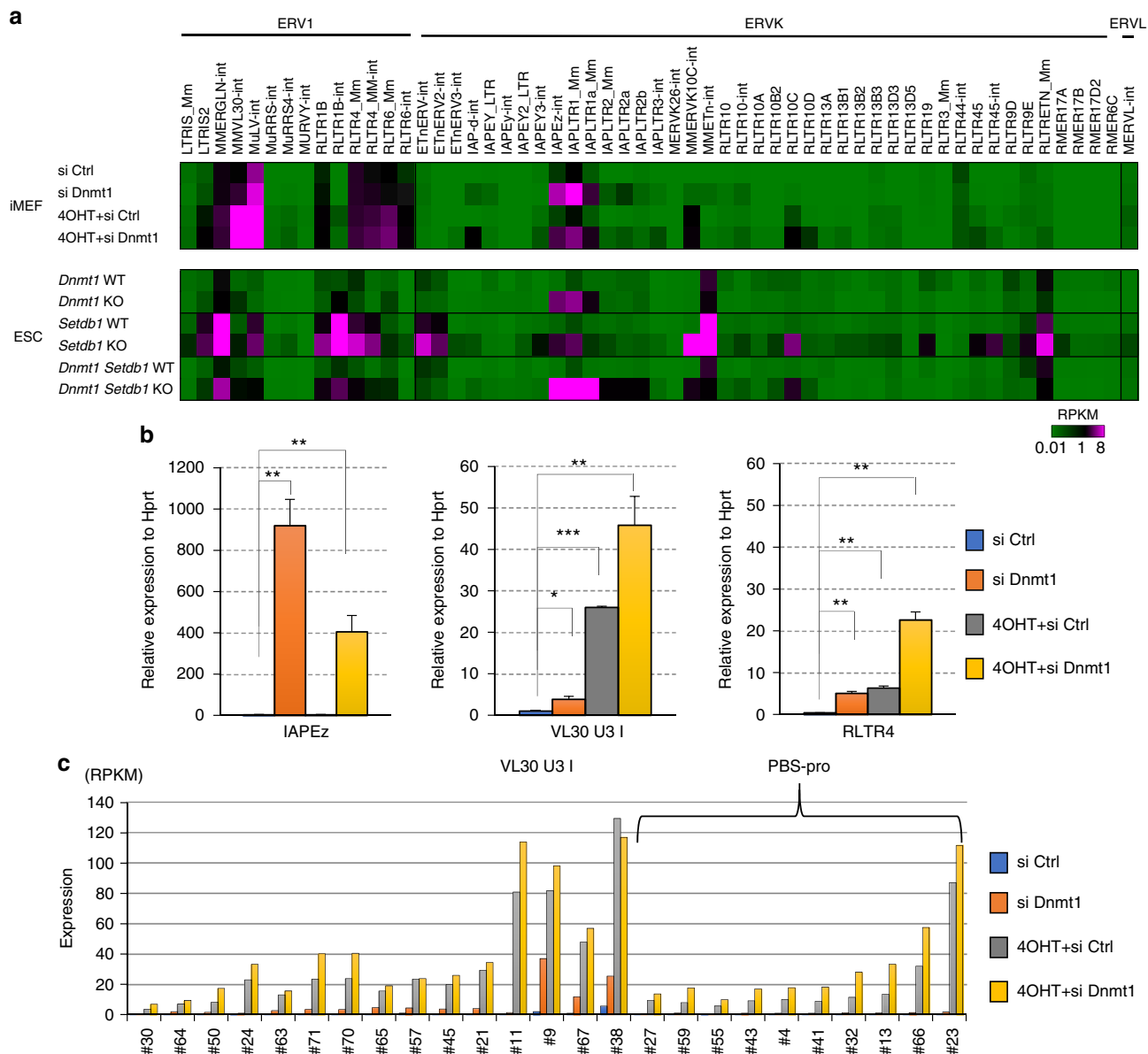
**Fig. 3** VL30 activation requires cell-type-specific TFs. **a, b** RT-qPCR of the VL30 U3 classes in RNA from *Setdb1* cKO iMEFs (**a**) or ESCs (33#6,<sup>3</sup>) (**b**), untreated or treated with 1  $\mu$ M all-trans retinoic acid (atRA), with or without 4OHT ( $n = 3$  biological replicates). For ESC data ( $n = 2$  biological replicates), we normalized each expression to VL30 U3 class I LTR expression of control iMEF. Error bars represent s.d. \*0.005 <  $P$  < 0.05, \*\*0.0005 <  $P$  < 0.005, \*\*\* $P$  < 0.0005, Student's  $t$ -test. **c** Upper panel; the derepression induced by *Setdb1* depletion in each VL30 U3 class I element (totally 71 elements) is shown. 22 loci on the right side have typical PBS-pro near the 3' end of 5' LTR. Selected loci for parallel alignment shown in **d** are denoted by H, M, and L symbols based on their derepression (H: high assigned to the locus where RPKM<sup>KO</sup> (RPKM of *Setdb1* KO) is >100, M: mild assigned to 10  $\leq$  RPKM<sup>KO</sup> < 100, and L: low assigned to RPKM<sup>KO</sup> < 10, respectively). Lower panel: H3K9me3 ChIP-seq read counts on 5' LTR at each VL30 U3 class I locus in *Setdb1* cKO iMEFs (no treatment (WT: blue bar) and day 7 after treatment of 4OHT (KO: light blue bar)) are shown. **d** Alignment of U3 sequences of VL30 U3 class I. The red box indicates the consensus sequence of Elk family binding sites. The blue and green boxes indicate the consensus sequences of AP-1 and Ets, respectively. High, mild, and low indicate in Fig. 3c upper panel. Entire alignment of VL30 class I shown in Fig. 3c is in Supplementary Data 3. We defined the copies for RPKM<sup>KO</sup> < 10 as silenced. **e** Inhibition of the MAPK pathway precludes the transcriptional activation of VL30 in *Setdb1* KO iMEFs. RT-qPCR analysis of VL30 U3 class I. 4OHT treatment was applied for the first 4 days and cells were harvested at day 6. The MEK inhibitor PD0325901 (PD; 1  $\mu$ M) was added for the last 24 h ( $n = 2$  biological replicates). Error bars represent s.d. \*0.005 <  $P$  < 0.05, \*\*0.0005 <  $P$  < 0.005, Student's  $t$ -test



*Setdb1* KO MEFs, but not in *Zfp809* KO MEFs (Fig. 5a). In the PBS-pro VL30 group, both *Setdb1* and *Zfp809* KO MEFs showed derepression of VL30 (Fig. 5a); however, derepression in the *Setdb1* KO was less prominent than that in the non-PBS-pro group, as shown in Fig. 3c. These results indicate that ERVs possessing non-PBS-pro escape repression by *Zfp809*, but are targeted by other silencing factors that can recruit *Setdb1*. *Zfp809* is required to initiate stable epigenetic silencing during development, but not to maintain silencing in somatic cells<sup>38</sup>.

To confirm whether *Trim28* is also required for the maintenance of ERV silencing, previously published RNA-seq data from *Trim28* KO MEFs, in which *Trim28* was depleted in

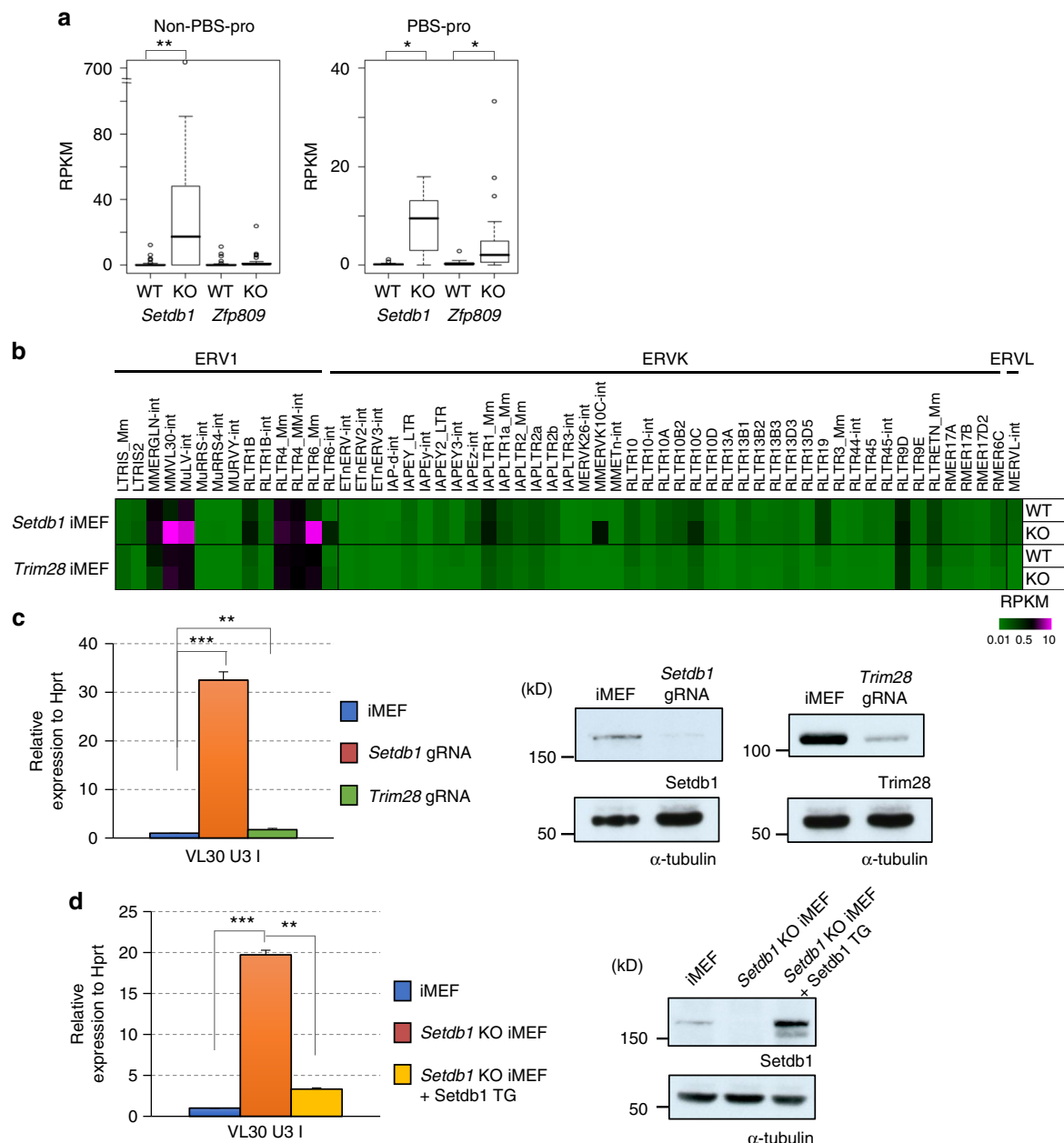
the conditional KO MEFs<sup>5</sup>, were compared with our *Setdb1* KO iMEF data. In addition to the lack of increase in IAPez expression, which was published previously, an increase of less than 2-fold in VL30 (MMVL30-int) expression in *Trim28* KO MEFs was observed (Fig. 5b). It is possible that the derepression of ERVs is desensitized after long-term culture of *Trim28* KO MEFs. Therefore, we utilized CRISPR-gRNA systems to inactivate *Trim28*, and analyzed the silencing of VL30 in iMEFs in a short time period. We also used the CRISPR-gRNA system for *Setdb1* inactivation. Both *Trim28* and *Setdb1* gRNAs efficiently diminished protein expression of their target molecules, although the cells were mixture of WT, and the incomplete and complete KO populations



**Fig. 4** Dnmt1 and *Setdb1* are required for the silencing of distinct sets of ERVs. **a** Expression of ERV families by the loss of *Dnmt1*, *Setdb1*, and simultaneous depletion of *Dnmt1* and *Setdb1* in iMEFs and ESCs. To obtain mRNA-seq data of iMEFs, *Setdb1* cKO iMEFs were transfected with siRNA against *Dnmt1* or control siRNA in combination with or without *Setdb1* depletion by the 4OHT treatment, then mRNA was isolated. RNA-seq data from *Dnmt1* cKO or double cKO of *Dnmt1* and *Setdb1* ESCs (6 days after treatment of 4OHT or without treatment)<sup>36</sup> were reanalyzed. Heatmap indicates the magnitude of the RPKM value. **b** *Setdb1* cKO iMEFs were transfected with siRNA against *Dnmt1* either alone or in combination with the loss of *Setdb1*. RT-qPCR of VL30 U3 class I RLTR6, IAPez, and RLTR4 was performed. Values represent mean expression relative to VL30 U3 I control ( $n = 2$  biological replicates). Error bars represent s.d. \* $0.005 < P < 0.05$ , \*\* $0.0005 < P < 0.005$ , \*\*\* $P < 0.0005$ , Student's *t*-test. **c** The derepression induced by *Dnmt1* KD, *Setdb1* KO, and simultaneous depletion of *Dnmt1* and *Setdb1* in iMEFs in each VL30 U3 class I element (totally 71 elements) is shown. The number of RNA-seq reads overlapping with a locus was divided by the length of the locus and normalized by million mapped reads (RPKM) as shown in Fig. 3c top

(Fig. 5c, right panel). The induction of U3 class I VL30 expression in iMEFs by this *Setdb1* gRNA treatment was greater than 30-fold induction. In contrast, we observed an ~2-fold increase in VL30 expression using *Trim28* gRNA (Fig. 5c, left panel), consistent with the results for the *Trim28* KO MEFs (Fig. 5b). Thus, these results indicate that *Trim28* is mostly dispensable for the maintenance of VL30 silencing in MEFs, like *Zfp809*.

Finally, to test whether restoring *Setdb1* expression can reverse U3 class I VL30 expression in long-term-cultured *Setdb1* KO iMEFs, we stably transfected cells with a transposon-based *Setdb1* expression vector. Exogenous *Setdb1* was capable of repressing VL30 upregulation, but this expression was still higher (~3-fold) than that in WT iMEFs (Fig. 5d). These results suggest that *Setdb1* is recruited to VL30 loci via additional, unknown factors to initiate silencing in iMEFs.



**Fig. 5** Distinct requirement of *Setdb1*, *Trim28*, and *Zfp809* for VL30 silencing. **a** Boxplots indicate silencing effects on VL30 U3 I by *Setdb1* and *Zfp809*. 71 VL30 U3 I (MMVL30-int with RLTR6\_Mm U3 I) are divided into two groups according to with (22 loci) or without (49 loci) PBS-pro sequences. RNA-seq data of *Setdb1* cKO iMEFs (no treatment (WT) or 4OHT 5d (KO)), *Zfp809* KO MEFs derived from mutant embryos (KO) and wild type (WT)<sup>38</sup> are shown. \*0.005 < P < 0.05, \*\*0.0005 < P < 0.005, Student's t-test **b** Expression of ERV families in *Trim28* KO MEFs. RNA-seq data for *Trim28* WT and KO MEFs<sup>5</sup> was reanalyzed. Heatmap indicates the magnitude of the RPKM value. For comparison, *Setdb1* cKO iMEF data (Fig. 1a) was shown above. **c** *Trim28* KO only marginally lead to VL30 derepression. iMEFs were transfected with a CRISPR-gRNA vector against *Setdb1* or *Trim28*. Two days later, transfected cells (tRFP-positive) were sorted using FACS Aria. Four days later, cells were harvested, and RT-qPCR of VL30 U3 class I was performed. (*Trim28*; n = 3 biological replicates *Setdb1*; n = 3 technical replicates) Error bars represent s.d. \*\*0.0005 < P < 0.005, \*\*\*P < 0.0005, Student's t-test. Efficiencies of CRISPR KO were examined by immunoblotting with anti-*Setdb1* or anti-*Trim28* antibodies. **d** Reintroduced *Setdb1* represses VL30 in *Setdb1* KO iMEFs. *Setdb1* was stably expressed in *Setdb1* long-term-cultured KO iMEFs. RT-qPCR of VL30 U3 class I was performed. (n = 2 biological replicates). Error bars represent s.d. \*\*0.0005 < P < 0.005, \*\*\*P < 0.0005, Student's t-test. Expression level of *Setdb1* was examined by immunoblotting

## Discussion

In this study, we found that *Setdb1* has a constant role in ERV silencing, which serves as a layer of epigenetic silencing not only in early embryonic or germ-lineage cells, but also in further differentiated somatic cells, even though a subset of cell-type-specific ERVs are only derepressed in *Setdb1*-deficient settings. Other transcriptional mechanisms, including cell-type-specific TFs, may restrict the expression of ERVs in *Setdb1* KO cells (Fig. 6).

As described, distinct ERV families were derepressed in different cell types when *Setdb1* was depleted (Fig. 1a). However, the enrichment of H3K9me3 on ERVs targeted and repressed by *Setdb1* in mESCs was mostly maintained in various differentiated somatic cells. This situation was similar to that of VL30, whose expression was highly induced in iMEFs and GMPs, but not in ESCs, after *Setdb1* depletion. These results suggest that *Setdb1* continuously targets similar ERV classes and deposits H3K9me3 marks, regardless of the developmental stage or cell type and the expression level of *Setdb1*. It is then necessary to determine how different types of ERVs are activated in different cell types after *Setdb1* depletion. Two distinct mechanisms are probably critical. According to one mechanism, ERVs are repressed by multiple epigenetic pathways or marks, and they may function redundantly with *Setdb1* to silence some ERVs. *Setdb1*-mediated H3K9 methylation and *Dnmt*-mediated DNA methylation is an example of this mechanism. IAPez is redundantly repressed by *Setdb1* and DNA methylation in mESCs<sup>4</sup>. In addition to DNA and H3K9 methyltransferases, many other epigenetic or chromatin factors, such as *Lsd1/Kdm1*, the polycomb repressive complex 2, Yin yang 1 (*Yy1*), *Erb3-binding protein 1* (*Ebp1*), and *CAF-1*, have roles in ERV silencing<sup>42–46</sup>. Therefore, it is possible that these factors and deposited epigenetic marks may exhibit crosstalk and interact to cause ERV silencing in a context-dependent manner. In the second mechanism, tissue-type-specific or cell-type-specific TFs restrict the transcriptional competence of each ERV. Previously, we found that MLV-type ERVs are specifically derepressed in B cells after *Setdb1* depletion<sup>19</sup>. In this case, the expression of B-cell lineage-specific TF *Pax5* confers competence for such an MLV

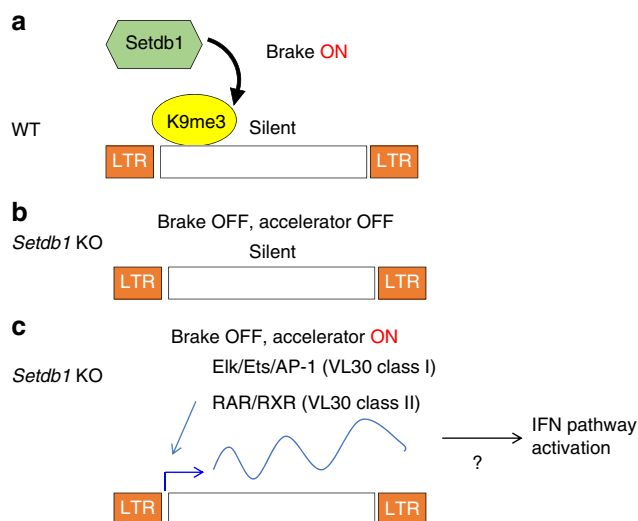
transcription. As shown in the current analyses, the cell-type-specific derepression of VL30 is also regulated by multiple cell-type-specific TF signaling pathways (Fig. 3). Thus, quiescent *Setdb1*-targeted ERVs have the potential for derepression upon inactivation of *Setdb1*, if their transcriptional machineries are competent.

Although it is generally recognized that DNA methylation, and not *Setdb1*-mediated H3K9me3, dominantly contributes to or is essential for ERV silencing in differentiated embryonic or adult somatic cells, only limited classes of ERVs, mostly IAPez and its flanking LTR, *IAPLTR1\_Mm*, are derepressed in *Dnmt1* KD or KO iMEFs. Since IAPez is stably and highly methylated, and this hyper DNA methylation is maintained in *Setdb1* KO ESCs<sup>3</sup>, it is reasonable to assume that IAPez is not derepressed by *Setdb1* KO alone in iMEFs. IAPez is a high-copy-number ERV in the mouse genome (~5000 copies, either full length or internally deleted), compared to members from other ERV families. Thus, the regulation of IAP was somewhat misleading with respect to the general behavior of ERVs.

In MEFs, only 1% (two genes) of gene promoters upregulated in *Setdb1* KO iMEFs were marked by H3K9me3 (Supplementary Fig. 3a), indicating that only a minority of induced genes are directly controlled by *Setdb1*. This finding is consistent with the results obtained for ESCs, but far fewer genes are regulated by *Setdb1* in MEFs<sup>4</sup>. The majority of upregulated genes in *Setdb1* KO MEFs were IFN pathway-related genes. A recent work by Cuellar et al. showed that the loss of *Setdb1* induces the derepression of retrotransposable elements, and the generated double-stranded RNAs activate the cytosolic RNA-sensing IFN signaling pathway<sup>47</sup>. The depletion of *Dnmt1* using RNAi or 5-Aza-dC treatment also induced IFN pathway-related genes (Supplementary Fig. 10). Thus, these upregulated genes are most likely indirectly induced by the derepression of ERVs.

Other H3K9me3-specific methyltransferases are *Suv39h1/2*. *Suv39hs* are responsible for H3K9me3 deposition at the pericentric heterochromatin, containing major satellite repeats. They also deposit H3K9me3 at intergenic major satellite repeats, but only if intact consensus repeat sequences are maintained<sup>24,48</sup>. Previous genome-wide analyses in ESCs, NPCs, and iMEFs have not provided clear evidence of a major function of *Suv39h*-dependent H3K9me3 in directing gene transcription<sup>24</sup>. Thus, a main function of H3K9me3 marks is the repression of retro-elements, such as exogenous retroviruses, ERVs, and LINE elements, or imprinted genes, which are also thought to be derived from retrotransposons, but not coding genes. This is consistent with the role of H3K9 methylation in other species, such as fission yeast and plants<sup>49,50</sup>.

ERVs are repressed by distinct, co-operative epigenetic mechanisms during the first few days of embryogenesis<sup>15</sup>. KRAB-ZFP family members are implicated in this process. KRAB-ZFPs bind to specific ERV loci and recruit their cofactors Trim28, *Setdb1*, and other silencing factors<sup>15</sup>. *Zfp809*, which belongs to the KRAB-ZFP family, represses VL30 by binding to PBS-pro, which is used by some retroviruses to prime reverse transcription<sup>38</sup>. *Zfp809* is required to initiate ERV silencing during embryonic development, but becomes largely dispensable in differentiated somatic tissues. *Zfp809* functions to repress only ERVs with the PBS-pro sequence. Thus, *Zfp809* regulates small fractions of VL30 repressed by *Setdb1* (Fig. 5a). For the silencing of ERVs without the PBS-pro sequence, other KRAB-ZFPs may bind to several loci of ERVs<sup>51</sup>. *Trim28* KO MEFs or acute depletion of *Trim28* in iMEFs leads to the marginal derepression of VL30 (Figs. 5b, c). Thus, the KRAB-ZFPs-Trim28 system is mostly dispensable for the maintenance of *Setdb1*-mediated VL30 silencing in MEFs. Lastly, it has been reported that a few number of VL30 elements were derepressed by treatment with a



**Fig. 6** Model of *Setdb1* function in ERV silencing. **a** *Setdb1* deposits H3K9me3 at ERV loci generally in differentiated cells, not only in ESCs. **b** In *Setdb1* KO cells H3K9me3 is reduced at ERV loci. However, TFs are required for each ERVs to get depressed. Without accelerator (TFs) ERVs are not expressed even without brake (H3K9me3). **c** For example, in *Setdb1* KO iMEFs derepression of VL30 class I require Elk, Ets, and AP-1 TFs. VL30 class II is expressed only in the presence of RA in *Setdb1* KO iMEFs. This ERV derepression might cause IFN pathway activation<sup>47</sup>

HDAC inhibitor, trichostatin A (TSA)<sup>52</sup>. We also performed an RNA-seq analysis of TSA-treated iMEFs, and could reproduce the reported findings. Interestingly, the impact of TSA was dominantly on the non-PBS-pro type elements (Supplementary Fig. 11). Thus, there must be more complex mechanisms for VL30 silencing. Future screen-based studies should be useful to identify genes or epigenetic factors responsible for Setdb1-mediated VL30 silencing in *Trim28* KO iMEFs.

## Methods

**Mouse embryonic fibroblasts.** *Setdb1* cKO iMEFs were described earlier<sup>3</sup>. Single cells were recloned by serial dilution. *Setdb1* deletion was induced by 800 nM 4OHT treatment for 4 days and further cultured without 4OHT for 2 days. *Setdb1* expression was determined by western blot analysis to screen clones that responded well (MEF3-12). Anti-*Setdb1* antibody (Cell Applications Inc., CP10377; dilution: 1:1,000) was used. To generate long-term-cultured *Setdb1* KO iMEFs, 800 nM 4OHT was added to MEF3-12 for 4 days, and further cultured without 4OHT for more than 1 week. For the CRISPR-gRNA knockout, pL-CRISPR-EFS-trFP plasmids<sup>53</sup> were transfected into iMEFs (MEF3-12) using the MEF2 Nucleofector Kit (Lonza). trFP-positive cells were sorted using FACS Aria. The efficiency of knock out was confirmed by a western blot analysis with anti-*Trim28* antibody (Abcam, ab22553; dilution: 1:1,000). Anti- $\alpha$  tubulin antibody (Sigma, T5168; dilution: 1:5,000) was used as an internal control. For chemical treatments, iMEFs were treated with 1  $\mu$ M atRA (Sigma) for last 2 days, 1  $\mu$ M MEK inhibitor PD0325901 (Wako) for last 24 hr, 5  $\mu$ M Ezh2 inhibitor GSK126 (Xcess Bio) for last 5 days, and 150 nM TSA for 24 hr before cells were harvested. For *Setdb1* rescue experiment in *Setdb1* KO iMEFs pPB-*Setdb1*-IRES Bsd plasmid with pPBase plasmid was transfected. Cells were selected with 8  $\mu$ g ml<sup>-1</sup> Blastcidin (InvivoGen). To generate 3 $\times$  Flag tagged *Ets1* expressing *Setdb1* cKO iMEFs, pPB-3xFlag-*Ets1*-IRES puro plasmid with pPBase plasmid was transfected. Cells were selected with 1.5  $\mu$ g ml<sup>-1</sup> puromycin (InvivoGen).

**Western blot analysis.** Cells were lysed in RIPA buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5 % NP-40, 0.5 % Na-deoxycholate, and 0.1 % SDS) supplemented with protease inhibitor cocktails (100 $\times$ ) (Nacalai) and 0.1 mM PMSF (Nacalai). Western blot analyses were performed using the antibodies as indicated. Uncropped scans of western blot images are available in Supplementary Fig. 12.

**RNA extraction and quantitative real-time RT-PCR.** RNA was isolated using RNeasy Plus Mini Kit (Qiagen). DNase I (NEB) was added to the column. Total RNA (1  $\mu$ g) was used to generate cDNA, using the Omniscript RT Kit (Qiagen) and random primers. qPCR was carried out using Power SYBR Green PCR Master Mix (ABI) on the ABI StepOnePlus. The signals were normalized against *Hprt* signals. Primer sequences are shown in Supplementary Table 1.

**RNA-seq.** Approximately 54 to 73 million high-quality 75-bp paired-end reads per sample for the forebrain<sup>18</sup>, ~110 million 100-bp paired-end reads per sample for iMEFs (Ctrl, 4OHT 5d and long-term cultured), 51-bp single-read per sample for iMEFs (siCtrl, siDnmt1, 4OHT 6d and 4OHT siDnmt1), and ~15 to 22 million 60-bp single-read for GMPs were mapped to the mouse genome (mm9) using the TopHat2 splice junction mapper (version 2.0.12)<sup>54</sup> with parameter (-g 1) telling the software to report best alignment only once for multi-hit reads. Mm9 Refseq gtf file was downloaded from UCSC (<http://genome.ucsc.edu>) in Oct 22, 2014 and applied as gene annotation file. After obtaining the aligned bam files, the Cufflinks algorithm (version 2.2.1)<sup>55</sup> was used to calculate FPKM with the Refseq gtf file. Repeat elements were downloaded from the University of California at Santa Cruz RepeatMasker track (mm9) and we selected ERVs annotated with high SW (Smith-Waterman) probability scores ( $\geq 2000$ ). Sequences of RepeatMasker annotated ERVs were compared with Repbase consensus sequences and scored for percentage of similarity or length mismatch. We applied SW 2000 as a cutoff because sufficient LTR only types (such as IAPLTR1a, RLTR6\_Mm, or RLTR10C) could not be collected with high SW ( $>3000$ ). Differential expression was computed using BEDTools for ERVs (<http://bedtools.googlecode.com>) with a minimum of 1-bp overlap. According to published data<sup>28</sup>, MMVL30 subclasses were assigned to their neighboring RLTR6 repeats and reconfirmed the position from which each LTR starts by alignment with consensus sequence. The genomic positions of ERVs used for the analysis are shown in Supplementary Data 4.

**Native ChIP and crosslinked ChIP.** Native ChIP assays were performed as described previously<sup>3</sup>. A mouse monoclonal antibody against H3K9me3 (2F3) was used<sup>56</sup>.

The antibody was incubated with anti-mouse IgG Dynabeads (Veritas for 1 h on ice and overnight after the addition of MNase-digested chromatin. For crosslinked ChIP,  $1 \times 10^7$  cells were crosslinked with 1% formaldehyde at 25 °C for 10 min. Chromatin was extracted and then sonicated using Bioruptor USD-250 to obtain an average fragment size of 300–500 bp. Immunoprecipitation was performed

using Dynabeads with antibodies, followed by purification using QIAquick PCR Purification Kit (Qiagen).

**Bisulfite sequencing.** Genomic DNA was purified from MEFs and bisulfite-converted using MethylCode Bisulfite Conversion Kit (Thermo). PCR products were amplified using TaKaRa EpiTaq HS (TaKaRa) and subcloned using TOPO Cloning Kit (Thermo) for sequencing. CpG methylation was analyzed using the QUMA tool (<http://quma.cdb.riken.jp>).

**siRNA.** For knockdown experiments, 50 nM siRNAs targeting *Dnmt1* (siGENOME SMARTpool) or control siRNA (Dharmacon) was transfected into *Setdb1* cKO iMEFs using RNAiMAX (Thermo). Transfected cells were passaged 24 h after transfection. A second transfection was conducted with the same reagents on the following day. Cells were harvested at day 6 and RNA was isolated. For simultaneous KO of *Setdb1*, 4OHT was added at day 0.

**ChIP-seq.** For ChIP-seq of H3K9me3, a polyclonal antibody against H3K9me3 (abcam ab8898) was used. For library preparation, KAPA Hyper Prep kit (KAPA biosystems) was used. Approximately 200 million 50-bp paired-end reads for forebrain cells, and ~20 to 23 million 76-bp single-end reads for iMEF were mapped to the mouse genome (mm9) using the Bowtie2 short read aligner (version 2.2.3) with default parameters<sup>57</sup>. A multi-hit read was assigned to one site randomly selected from among valid alignments and duplicate reads were removed using Picard tools (<https://broadinstitute.github.io/picard/>).

**Public sequencing data.** Raw reads were downloaded from publicly available ChIP-seq (H3K9me3 in ESC: GSM1375155, GSM727425), (H3K9me3 in MEF: GSM1375168, GSM1375173), (H3K9me3 in GMP: DRX021712, DRX021713, DRX021716, DRX021717), and RNA-seq (*Setdb1* KO ESC established from conditional KO ESC: GSM727423, GSM727424), (*Dnmt1* cKO ESC and *Dnmt1*, *Setdb1* double cKO ESC: GSE77781), (*Dnmt1* WT, *p53*<sup>-/-</sup> and *Dnmt1*<sup>-/-</sup>, *p53*<sup>-/-</sup> iMEF: GSM1089793, GSM1089794), (*Zfp809* KO MEF established from KO embryo at E12.5: SRX487521), (*Trim28* KO MEF established from *Trim28* conditional KO MEF: GSM1916177, GSM1916178, GSM1916179, GSM1916180). Mapping and data processing were performed as described above.

**ChIP-seq NGS plots.** NGS plots show the average ratios of normalized read density between ChIP and input samples every 10 bp (1 bin) of H3K9me3 from -5 kb to 10 kb around genomic ERV elements in ESC, forebrain, iMEF, and GMP. We selected ERVs containing -int element (which means the internal region) with flanked LTRs. Position 1 is 5' start site of the -int element and average length of flanking LTR elements is about 600 bp. Positions of LTRs and int for each ERV element are indicated below the plots. Unique ERVs containing MMERVK10C-int (167 copies, average length: 7190 bp, flanking LTRs are RLTR10C), IAPez-int (593 copies, average length: 5516 bp, flanking LTRs are IAPLTR1a), RLTR4\_Mm-int (50 copies, average length: 7230 bp, flanking LTRs are RLTR4\_MM), RLTR6-int (186 copies, average length: 7212 bp, flanking LTRs are RLTR6\_MM), and MMVL30-int (86 copies, average length: 5294 bp, flanking LTRs are RLTR6\_MM) were analyzed. ChIP-seq data sets for ESC and iMEF<sup>24</sup>, GMP<sup>20</sup> and forebrain cells (this study) were used. Alignment files for H3K9me3 ChIP-seq data (bam format) were transformed to read coverage files (bigwig format) and processed to obtain plots for each ERV of interest using DeepTools<sup>58</sup> with the following parameters: -ratio log2, bin Size 10, scaleFactors SES. Read densities are normalized with read depth and million mapped reads of each NGS data.

**Data availability.** All NGS sequencing data that support the findings of this study are available in the NCBI Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) database under the accession number GSE102490.

Received: 6 October 2017 Accepted: 5 April 2018

Published online: 27 April 2018

## References

- Waterston, R. H. et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Goodier, J. L. Restricting retrotransposons: a review. *Mob. DNA* **7**, 16 (2016).
- Matsui, T. et al. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931 (2010).
- Karimi, M. M. et al. DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* **8**, 676–687 (2011).
- Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
- Thomas, J. H. & Schneider, S. Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* **21**, 1800–1812 (2011).



7. Wolf, D. & Goff, S. P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**, 1201–1204 (2009).
8. Liu, S. et al. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* **28**, 2041–2055 (2014).
9. Kobayashi, H. et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
10. Seisenberger, S. et al. The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
11. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96–99 (2004).
12. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**, 116–117 (1998).
13. Ellis, J., Hotta, A. & Rastegar, M. Retrovirus silencing by an epigenetic TRIM. *Cell* **131**, 13–14 (2007).
14. Wiznerowicz, M. et al. The Kruppel-associated box repressor domain can trigger de novo promoter methylation during mouse early embryogenesis. *J. Biol. Chem.* **282**, 34535–34541 (2007).
15. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411**, 273–287 (2011).
16. Quenneville, S. et al. The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. *Cell Rep.* **2**, 766–773 (2012).
17. Rowe, H. M. et al. De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519–529 (2013).
18. Tan, S. L. et al. Essential roles of the histone methyltransferase ESET in the epigenetic control of neural progenitor cells during development. *Development* **139**, 3806–3816 (2012).
19. Collins, P. L., Kyle, K. E., Egawa, T., Shinkai, Y. & Oltz, E. M. The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc. Natl Acad. Sci. USA* **112**, 8367–8372 (2015).
20. Koide, S. et al. Setdb1 maintains hematopoietic stem and progenitor cells by restricting the ectopic activation of nonhematopoietic genes. *Blood* **128**, 638–649 (2016).
21. Takikita, S. et al. A Histone methyltransferase ESET is critical for T cell development. *J. Immunol.* **197**, 2269–2279 (2016).
22. Fasching, L. et al. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep.* **10**, 20–28 (2015).
23. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
24. Bulut-Karslioglu, A. et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 (2014).
25. Pasquarella, A. et al. Retrotransposon derepression leads to activation of the unfolded protein response and apoptosis in pro-B cells. *Development* **143**, 1788–1799 (2016).
26. Nilsson, M. & Bohm, S. Inducible and cell type-specific expression of VL30 U3 subgroups correlate with their enhancer design. *J. Virol.* **68**, 276–288 (1994).
27. Islam, T. C. & Toftgard, R. Nuclear orphan receptor-binding retinoic acid response elements in keratinocytes. *Biochem. Biophys. Res. Commun.* **203**, 545–552 (1994).
28. Herquel, B. et al. Trim24-repressed VL30 retrotransposons regulate gene expression by producing noncoding RNA. *Nat. Struct. Mol. Biol.* **20**, 339–346 (2013).
29. Markopoulos, G. et al. Genomic analysis of mouse VL30 retrotransposons. *Mob. DNA* **7**, 10 (2016).
30. Janknecht, R., Ernst, W. H., Pingoud, V. & Nordheim, A. Activation of ternary complex factor Elk-1 by MAP kinases. *EMBO J.* **12**, 5097–5104 (1993).
31. Sebolt-Leopold, J. S. & Herrera, R. Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat. Rev. Cancer* **4**, 937–947 (2004).
32. Reddy, M. A., Langer, S. J., Colman, M. S. & Ostrowski, M. C. An enhancer element responsive to ras and fms signaling pathways is composed of two distinct nuclear factor binding sites. *Mol. Endocrinol.* **6**, 1051–1060 (1992).
33. Owen, R. D., Bortner, D. M. & Ostrowski, M. C. ras oncogene activation of a VL30 transcriptional element is linked to transformation. *Mol. Cell. Biol.* **10**, 1–9 (1990).
34. Hutnick, L. K., Huang, X., Loo, T. C., Ma, Z. & Fan, G. Repression of retrotransposon elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. *J. Biol. Chem.* **285**, 21082–21091 (2010).
35. Reddington, J. P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* **14**, R25 (2013).
36. Sharif, J. et al. Activation of endogenous retroviruses in Dnmt1(-/-) ESCs involves disruption of SETDB1-mediated repression by NP95 binding to hemimethylated DNA. *Cell Stem Cell* **19**, 81–94 (2016).
37. McCabe, M. T. et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108–112 (2012).
38. Wolf, G. et al. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* **29**, 538–554 (2015).
39. Sripathy, S. P., Stevens, J. & Schultz, D. C. The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Mol. Cell. Biol.* **26**, 8623–8638 (2006).
40. Wolf, D. & Goff, S. P. TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* **131**, 46–57 (2007).
41. Wolf, D. & Goff, S. P. Host restriction factors blocking retroviral replication. *Annu. Rev. Genet.* **42**, 143–163 (2008).
42. Schlesinger, S., Lee, A. H., Wang, G. Z., Green, L. & Goff, S. P. Proviral silencing in embryonic cells is regulated by Yin Yang 1. *Cell Rep.* **4**, 50–58 (2013).
43. Yang, B. X. et al. Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* **163**, 230–245 (2015).
44. Macfarlan, T. S. et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
45. Leeb, M. et al. Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* **24**, 265–276 (2010).
46. Wang, G. Z., Wolf, D. & Goff, S. P. EBP1, a novel host factor involved in primer binding site-dependent restriction of moloney murine leukemia virus in embryonic cells. *J. Virol.* **88**, 1825–1829 (2014).
47. Cuellar, T. L. et al. Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *J. Cell Biol.* **216**, 3535–3549 (2017).
48. Martens, J. H. et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800–812 (2005).
49. Lippman, Z. et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476 (2004).
50. Cam, H. P. et al. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.* **37**, 809–819 (2005).
51. Ecco, G. et al. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev. Cell* **36**, 611–623 (2016).
52. Brunmeir, R. et al. Epigenetic regulation of a murine retrotransposon by a dual histone modification mark. *PLoS Genet.* **6**, e1000927 (2010).
53. Heckl, D. et al. Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32**, 941–946 (2014).
54. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
55. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
56. Chandra, T. et al. Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell* **47**, 203–214 (2012).
57. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
58. Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

## Acknowledgements

We thank Ms. Chikako Shimura for her technical assistance with the western blot analysis and all other members of the Shinkai lab for their experimental support, critical feedback, and suggestions. We also thank Mr. Kenji Ohtawa (RIKEN BSI Research Resources Center: RRC) who provided technical support for cell sorting and Drs. Eugene M. Oltz, Patrick L. Collins, and Matthew C. Lorincz for critically reading the manuscript. Illumina sequencing was supported by Genome Resource and Analysis Unit RIKEN CDB. This work was supported in part by AMED-CREST and a RIKEN internal research fund.

## Author contributions

M.K. designed and conducted the experiments, K.T. designed the experiments and performed informatics analysis, and Y.S. designed and supervised the experiment. M.K., K.T., and Y.S. wrote the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-018-04132-9>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018